SAND2014-18342R

SAND 2014-XXX

Annual Report

Feasibility and Demonstration of a Cloud-Based RIID Analysis System

Project numbers:
SL13-FY13-187-PD03
OR13-CldBasedRIIDS-PD03

Lab: Sandia National Laboratories

Project PI and primary contributor: William Johnson

Date: 20140930

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Feasibility and Demonstration of a Cloud-Based RIID Analysis System

Project numbers: SL13-FY13-187-PD03, OR13-CldBasedRIIDS-PD03

1. INTRODUCTION

For fiscal year 2014 this project aimed to look at, characterize, and attempt to offer solutions to a few practical issues a cloud-based RIID system would have to deal with, namely:

- Automated check of detector energy calibration
- Aging effects of deployed detectors
- Detector identification from submitted spectra when no meta-information is available

The NA-42 Triage dataset was determined to be a best source of data to perform these studies with, since this data was all taken and submitted in the field, both from training exercises, as well as real world events. Additionally, the data available from the Triage website spanned from about June of 2005 to present, which provides some possibility of finding data submitted by the same deployed detector, over the time span of years. The bulk of the work performed this year was in retrieving, organizing, and obtaining the capability to extract the necessary information from the spectrum files, so the three listed studies could then be performed.

2. OBTAINING AND PARSING SPECTROSCOPIC DATA

There are about 35,000 data files available from the Triage web page, so manual retrieval and organization of the files was not an option. Instead, the website operators agreed, and got approval, to give us access to the internal API's to retrieve the files. However, this access required giving us a certificate to the Honeywell network, which approval for stalled out in the IT security department. To get around this, we instead used a headless web browser (a browser with no display), and wrote a script to download and organize the data and submission details from the Triage website. This script was non-trivial due to the dynamic nature of the website. This delay stalled the project until the third quarter of the year.

To extract both the spectroscopic and meta-data included in the files produced by the detectors, we ended up having to extend some file parsing code the author wrote for another project (the *InterSpec* web-based analysis tool). The other existing batch parsing solution, Cambio, did not preserve enough of the meta-information inside the files to be useful for this project; it also failed to parse a number of relevant file formats. Permission to use the *InterSpec* file parsing code was obtained, and it was extended to extract the necessary meta-information from many existing formats it parsed, as well as a number of new formats were added to its code so it now parses about 99% of the spectrum files, with the remaining files from detectors irrelevant or un-usable by this study. The Triage dataset also required extensive cleanup, since many of the files from the field where images, text files, corrupted, or otherwise irrelevant or un-usable:

- ~35,000 files retrieved from the Triage website
- 14,432 candidate spectroscopic files remained after filtering by file extension
- 14,067 files successfully parsed with spectroscopic information available
- 7,850 files contain an individual detector identifier, such as a serial number
- 2,659 of these could be matched to a submitter specified detector type

The date ranges and number of spectra received from some of the detectors that have submissions spanning the longest time periods are shown in Table 1. Some initial observations are that Ortec Detective detectors top this list (the 15 longest spanning detectors are Ortec Detectives, and 69 of the top 100), and that submitters often times misidentify the model of the detector (ex. specify Detective-EX100 instead of Detective-DX) about half the time. It is also worth noting that changes to the configurations (like energy range capability) of the detectors are common, but has not been quantified yet; these types of changes will also need to be detected and accounted for by a cloud-based RIID system.

• First Date	Last Date	# Spec.	Type
20051002	20130924	80	Some Detective
20061109	20130205	92	Some Detective
20071214	20140805	39	Some Detective
20070808	20140416	90	Some Detective
20080422	20140905	467	Some Detective
20070808	20131005	135	Some Detective
…	…	…	…
20091108	20130609	5	GR-135
…	…	…	…
20100729	20140522	62	Some identiFINDER
…	…	…	…
20050915	20090211	4	GR-135
20100306	20140502	60	Some Detective
…	…	…	…
20101006	20130910	35	Some identiFINDER
…	…	…	…
20080213	20110817	4	GR-135
20080725	20110914	5	GR-135

Table 1: The date range available of individual detectors with the largest ranges, as well the number of spectra submitted with the detector, and the model. The list is ordered according to the length of time span data has been received over, in decreasing order, for individual detectors. Rows with ellipses indicate omitted rows of Detective detectors. The Detective and identiFINDER systems are commonly incorrectly reported as the different sub-models (initial estimates indicate wrongly about 50% of the time), so the exact model is not given.

3. PEAK IDENTIFICATION AND FITTING

Surprisingly little could be found in the literature for the automated detection of peaks when the background, width, location, and amplitude are all unknown. The use of peak properties is central to all three studies to be performed for this project, so we decided to develop our own algorithm based off of the second-derivative of the spectroscopic data.

We initially tried a few novel methods not involving the second derivative, but their performance was equal or inferior to the second derivative method, so were not used.
It is well known the second-derivative of a Gaussian peak reaches its most negative value at the centroid of the peak, however decision criteria for finding peaks in low statistics regions, areas where multiple peaks are located, areas where other spectral features are present, with variable numbers of channels per peak, or other conditions commonly ran into, were not available in the literature, and none of the automated peak detection algorithms we tried in available software packages performed sufficiently.
The second derivative of a high-statistics, “clean” spectrum can be seen in Figure 1, as well as a more typical low-resolution spectrum and associated smoothed second derivative. As can be seen from the figure, peaks can be indicated from the second derivative, but false positives can also be common. There are also a number of circumstances were peaks are not indicated by the second derivative, such as in the low resolution spectrum of Figure 1, there are 4 dominant peaks in the 250 to 415 keV range, but the second derivative will indicate two at best. So in order to optimize peak detection, we decided to manually fit for peaks in a number of spectra from commonly used detector systems (about 200 spectra currently), and then optimize a number of variables in the peak searching algorithm against the manually identified peak properties. Then further work is performed using maximum likely hood fits to a continuum and Gaussian, followed by a further scan now that estimates can be made of the FWHM response of the detector. A full description of the algorithm is not given here since it is not fully developed yet, and is still being improved.

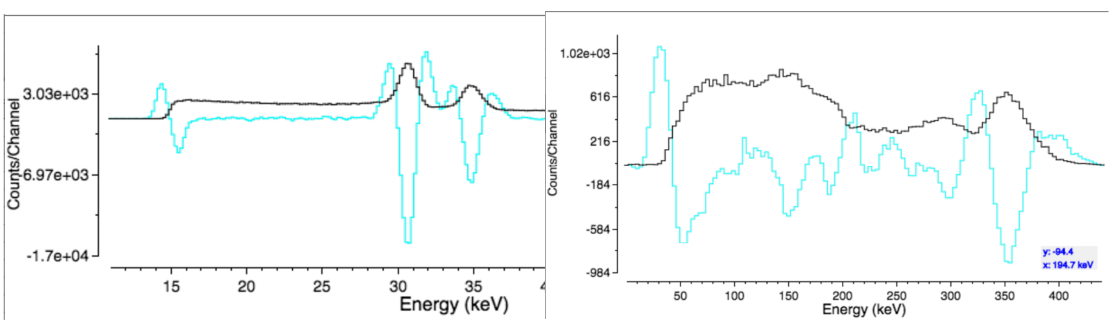

Figure 1: The Savitzky-Golay filtered second derivative (teal line) of a [Left] high-statistics HPGe “clean” spectrum (black line), and [Right] a low-resolution spectrum of Ba-133. The second derivatives have been scaled to be on a similar scale as the spectra.

4. PATH FORWARD

Once development of the peak search algorithm is complete, this information in combination with the other meta-information that has been extracted will be used to complete the automated check of energy calibration, and for automatic identification of detector model purely from the spectral information.
It was hoped that enough spectra from check sources, along with their distances and activities would have been submitted to help determine changes to intrinsic detector efficiencies over time, but an initial investigation of the data shows that this is not the

case; many of the events where this information was specified have turned out to be inject data, or of questionable integrity. We will look into this further, as well as use the peak search/fit algorithm to track detector FWHM response over time, and manually look for any more characteristics we can identify as being important to account for as detectors age.

5. PRESENTATIONS AND PUBLICATIONS

No presentations or publications were made in this fiscal year.